General Notes on Choosing Smart or Standard Script Variants:

For each of the 5 scripts in this suite, there are 2 variants present. One is titled ‘Smart’, and the other isn’t. Those that are titled ‘Smart’ will adapt to whatever data you throw at them, from any organism, with any chromosome configuration.

Those that are not, are hard-coded for wheat, its chromosomes, and the specific maps that were used as references. In most cases, using the Smart variant will be more convenient, adaptable, and require little if any modification. However, the manually determined parameters for each plot that are set in the standard script variants do potentially offer stability, control, and intercomparability between plots of different data sets. For example, in the Smart scripts, if all your markers present in your reference file are on the first 1/4th of a chromosome, the ‘Smart’ plots will scale to fit these, and the labeled axes will adjust accordingly. If, however, you had the axes fixed in place, it would be easier to see that these markers only represented part of one chromosome arm without having to reference the basepair and cM lengths of the chromosome in question. In most cases, this should not be an issue, but if it is, and a group wants to generate many easily cross-compared plots where the graph window represents a consistent size of the full chromosome, then it may be useful to use the non-smart versions of these scripts.

Smart\_Basic\_MAST and Basic\_MAST:

Function:

These scripts map all the markers across all chromosomes in a file by both cM and basepair position if the file is prepared correctly.

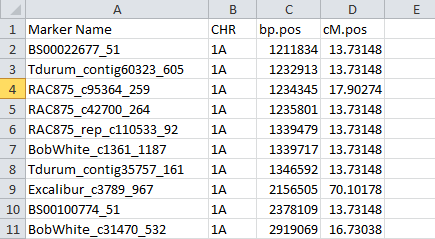
Requirements:

To use these scripts, you need 1 external .csv file. It needs to have at least 4 columns, containing respectively the marker names, the marker chromosome, the marker centimorgan position in a linkage/consensus map, and the marker physical position in basepairs.

Formatting:

The file should be formatted like the screenshot below, with the crucial headings being “CHR”, “bp.pos” (standing for basepair positions), and “cM.pos” (standing for centimorgan position).

The markers in this example file are ordered, but importantly, they do not have to be, the script will correctly subset the data and plot it regardless of order.



Smart vs. Manual:

The Basic\_MAST script plots out each chromosome individually, which can be useful under certain circumstances, but requires manual adaptation for non-wheat organisms.

The Smart\_Basic\_MAST script has a single loop that will plot out every

Running the scripts:

Open R or R Studio, ensure that the working directory and file names in the script coincide with the location and files you want to access, and run the code. Ensure that the read.csv command retains the “,header=TRUE” statement, as this is critical in denoting the vectors.

For the variable called OutputName, seen in the codes as:

OutputName<-("Smart\_Example\_Basic\_MAP")

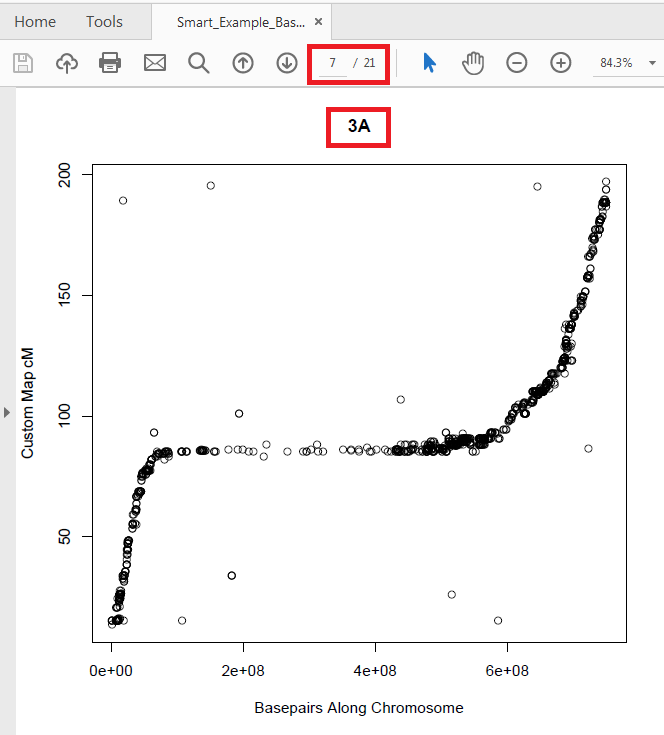
Set this to whatever name you want to give your analysis. You can leave it as this example name and it will still run, but if you’re creating multiple maps, you’ll want to give each a unique name.

Output:

Once the script has been run, look at your working directory again. A PDF file should have been created there, named by whatever you set the OutputName variable to be.

If you open this file, you’ll see that there are as many pages in it as there are chromosomes in your dataset.

For each chromosome within the dataset, the physical and linkage positions will be plotted out, each one labeled with the name



Smart\_Double\_MAST and Double\_MAST

Function:

These scripts plot 2 series of markers from two files, one over the other, in different colors, for every chromosome. While theoretically, any 2 series of can be plotted, generally this is most useful for highlighting a specific subset of markers, or comparing 2 maps.

Requirements:

To use these scripts, you need 2 external .csv files. Each containing a list of markers with chromosome data, cM positions and basepair positions, as with the file for Basic\_MAST.

Formatting:

The formatting of the .csv has the same requirements as Basic\_MAST, for both files.

For Smart\_Double\_MAST, the script should function without notable alterations for any organism or map, but for ordinary Double\_MAST, the boundaries of the graphs and the chromosome names will need to be manually changed to match your organism and map in question. For this script, it is vital to fix the X and Y limits of the graph in place, or the plotting of the second series will be scaled separately from the first.

Running the scripts:

This is the same as in Basic\_MAST.

Open R or R Studio, ensure that the working directory and file names in the script coincide with the location and files you want to access, and run the code. Ensure that the read.csv command retains the “,header=TRUE” statement, as this is critical in denoting the vectors.

UP\_MAST and Smart\_UP\_MAST

Acronym: Useful Polymorphic Markers Around Segregating Traits

Function:

These scripts identify polymorphic markers between any two individuals. Those polymorphic markers are then output into a separate .csv file that can be browsed. Then the scripts make chromosome plots, background markers are plotted by both physical position and linkage position, then the polymorphic markers between the 2 individuals are plotted on top of them in a separate color. Then on top of those plots, there is the optional plotting of labeled vertical lines, denoting the physical positions of specific loci of interest, and optionally labeling the genotypic state of each of these 2 individuals at that locus.

Requirements: 3 .csv files

File 1 = all markers on the chromosome with the same format as the Basic\_MAST file.

File 2 = A marker call database file, with marker names, marker chromosomes, marker basepair positions, marker cM position, and genotyping data from multiple (at least 2) genotypes/cultivars/accessions.

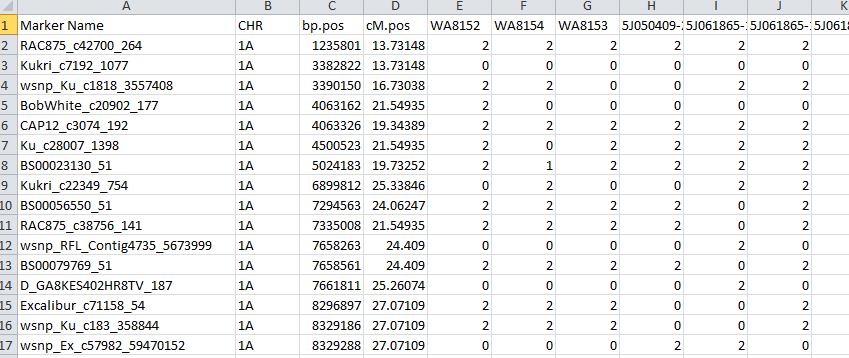
File 3 = a file that highlights specific loci.

These 3 files are technically independent, and this script can be run without file 1, or without file 2, or without file 3.

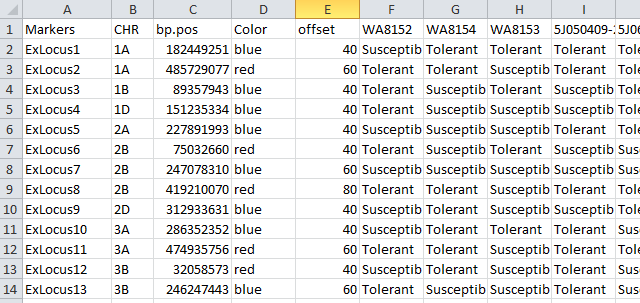
Formatting:

File 1 should be formatted as the files for Basic\_MAST and Double\_MAST were.

File 2 should have the first 4 columns formatted as File 1 is, but should then have the columns for each genotype. Genotyping values (the results of a given marker for a given genotype) should be in 0/1/2 format, where 0 and 2 represent homozygous opposing SNP/marker calls, and 1 represents a heterozygous marker call. See the example below.



File 3 (optional) needs marker/loci names, chromosome positions, basepair positions of the locus, a color that the locus will be marked in, an offset value (in cM, used for adjusting labels up and down for visibility on the graph in crowded circumstances), and optionally, genotyping data on the locus for any genotype you want to examine. The labels must be spelled consistently between the headings and the script. The presence of special characters in the names should not cause any problems.



File 3 does not include a cM column, and its input is not required to be (though can be) specific markers, it merely requires specific loci, though the column is labeled Markers.

Running the Scripts:

In the code, type in the names (exact spelling and capitalization) of two genotypes from your database in sheet 2 and/or sheet 3 where it is requested for Genotype1, and Genotype2.

Then ensure that the commands in the script are correctly referencing the working directory, and the 3 file names.

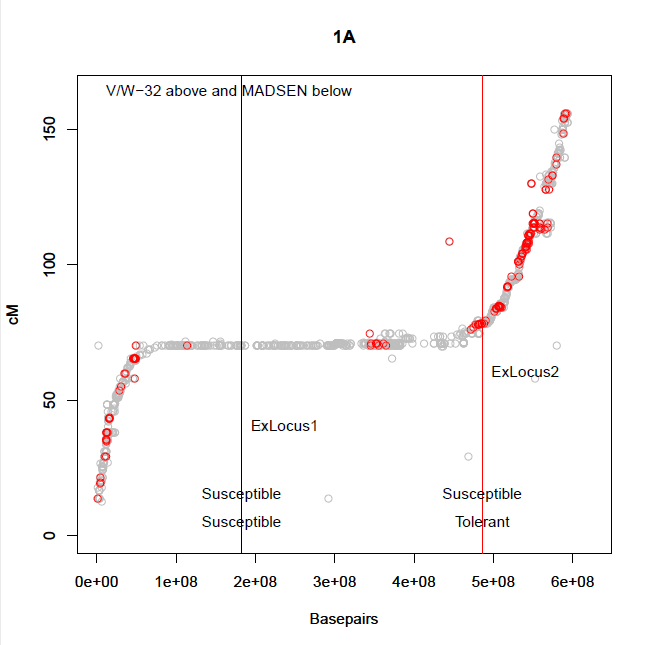
Run the script, the output should be in 2 files, a .csv of the polymorphic markers, and a .pdf of the plots. Both of those files will include the names of the 2 genotypes that were examined.

Some warnings will occur, where the script will have attempted to plot empty vectors, but these should not be concerning, and the data should be correctly plotted if the formatting was correct.

Output:

The .pdf output should resemble this, with 1 page for each chromosome.

In cases where the genotypes



CROS\_MAST and Smart\_CROS\_MAST

(Cross Referencing Other Sources - Markers Around Segregating Traits)

Function:

This script is a more complex version of UP\_MAST designed to deal with situations where parental data is missing.

Sometimes one would like to know what markers will be polymorphic in a given population, but doesn’t have genotyping data directly from the parental genotypes. This script allows the user to draw data from grandparents of the population as a proxy for either one or both parents.

Requirements: 3 .csv files (the same requirements as UP\_MAST)

File 1 = all markers on the chromosome with the same format as the Basic\_MAST file.

File 2 = A marker call database file, with marker names, marker chromosomes, marker basepair positions, marker cM position, and genotyping data from relevant parental/grandparental genotypes.

File 3 = a file that highlights specific loci.

Formatting: The same as UP\_MAST

Running the Scripts:

The basic structure and procedure of UP\_MAST remains intact for this script, with one critical difference at the front. Instead of simply having Genotype 1 and Genotype 2, either or both of those can be missing.

The section where you must enter in your data for the analysis reads:

“

GenotypeName1<-("")

GenotypeName2<-("")

GenotypeName11<-(" ")

GenotypeName12<-(" ")

GenotypeName21<-(" ")

GenotypeName22<-(" ")

AnalysisType<-3

“

Genotypes 1 and 2 represent the parents

Genotypes 11, 12, 21, and 22 represent the grandparents, or the parents of Genotype 1 and Genotype 2 respectively.

Importantly, if you only have 1 parent missing, for this version of this script, you must designate that missing genotype as genotype 1 for it to function properly, thus providing genotype names for genotypes 11 and 12. The script only pulls out the data from the genotypes coded to each analysis type.

The referenced indices are as follows:

Analysis type 1 (identical to UP\_MAST): Genotypes 1 and 2

Analysis type 2 (1 parent missing, using grandparents of that missing parent as proxies): Genotypes 11, 12 and 2.

Analysis type 3 (both parents missing, using all 4 grandparents as proxies): Genotypes 11, 12, 21 and 22.

Enter in the names of the genotypes in your Sheet 2 database into their appropriate spots, and set the AnalysisType variable to either 1, 2 or 3, to fit the analysis you want to run.

Caveat: This script will not work for back-crossed populations as it is modeled. This is because all of the monomorphic markers it pulls out of the grandparents will match the parent. If you have a back-crossed population, either entering in the 2 most recent genetic sources that you have data on as parents and using analysis type 1, or running a standard UP\_MAST with those same settings would deliver you a list of probably segregating markers. This would still be dependent upon the selfing generations in between the back-crossing.

Select\_MAST and Smart\_Select\_MAST

Function:

It is sometimes useful to be able to select subsets of quality markers that are evenly spaced along the chromosome, but which the data shows are consistently progressing along the chromosome in both cM and basepair positions. This script provides the user with a graphical interface assist with manual selection of optimal markers.

Requirements:

1-3 .csv files.

Sheet 1: Background chromosomal trace, plotted in light grey. Due to formatting, this file is required, but can be blank apart from the headers. A blank example file is provided.

Sheet 2: Absolutely essential sheet, these are the markers you’re going to be selecting from.

Sheet 3: Horizontal gridlines sheet, can be modified for any spacing or to highlight key regions.

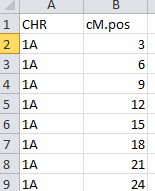
Formatting:

Sheet 1: Same as Basic\_MAST

Sheet 2:Same as Basic\_MAST or Double MAST

Sheet 3: This is a custom Sheet, designed to produce gridlines for each chromosome. It only requires 2 columns, one to designate the chromosome name, and the other to designate the cM position at which the user wants each grid line plotted. [There were more concise ways to code this, but I wanted flexibility.] A reference assembly sheet with excel commands to quickly produce an analogous sheet to this for a variety of situations is provided.

Example of Sheet 3



Running the Scripts:

This script relies heavily on the Base R function: Identify.

Select the code and hit run, and you’ll notice that the graphical window will plot your first chromosome. Hover your cursor over the plot, and you’ll notice that it turns into a large + sign. You’ll also notice that at the top right corner of the plot is a button that reads “Finish”.

The + cursor is an active interface that allows you to click on the plot. If you are sufficiently close to a marker plotted from Sheet 2, a small blue indicator sign will appear to indicate that you have selected that marker. You can continue selecting markers, and all of them will be recorded.

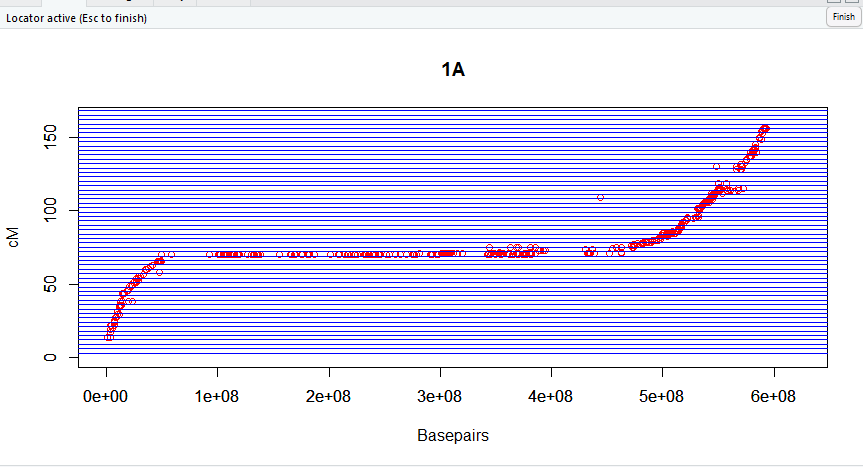


When you have selected all the markers you wanted from a given chromosome, click the finish button.

This should bring up the next chromosome plot, and you can repeat the process until you have selected all your desired markers for each chromosome. If you don’t want to select any markers on a specific chromosome, you can hit the finish button without clicking on any markers, and no markers from that chromosome will be added to the record. If you click on a place where no marker is present or near enough to your cursor, then a message will appear in your R console saying “warning: no point within (x) inches”. If you click, and this message does not appear, then you have selected a marker. If you click on the same marker more than once, you’ll get a message in the R console that says “warning: nearest point already identified”.

Once you have clicked finish on the last plot, you’ll notice that the plot becomes static, and your cursor no longer becomes a + sign if you move it over the plot. A .csv file with the output name you assigned should also now have appeared in your working directory with the full list of the markers you selected, their chromosomes, and positions.

Example of Smart\_Select\_MAST interface with a blank Sheet 1.



Output:

The first row after the heading is a place holder. All rows after that should be the markers you selected, listing the names, the chromosome you selected them from, the basepair position, and the centimorgan position of each from your data.

